On the Hardness of Approximating k-Dimensional Matching

E. Hazan * S. Safra * and O. Schwartz *

Abstract

We study bounded degree graph problems, mainly the problem of k-Dimensional Matching *(k-DM)*, namely, the problem of finding a maximal matching in a k-partite k-uniform balanced hyper-graph. We prove that k-DM cannot be efficiently approximated to within a factor of $O(\frac{k}{\ln k})$ unless $P = NP$. This improves the previous factor of $\frac{k}{2^{O(\sqrt{\ln k})}}$ by Trevisan [Tre01]. For low k values we prove NP-hardness factors of $\frac{54}{53} - \varepsilon$, $\frac{30}{29} - \varepsilon$ and $\frac{23}{22} - \varepsilon$ for 4-DM, 5-DM and 6-DM respectively. These results extend to the problem of Maximum Independent-Set in $(k+1)$-claw-free graphs and the problem of k-Set-Packing.

*Department of Computer Science, Tel Aviv University, Israel, {eladh,safra,odedsc}@post.tau.ac.il

1 Introduction

This paper studies two related combinatorial optimization problems, which are bounded variants of Independent-Set and Maximal-Matching in a hyper-graph. A hyper-graph is a set of vertices and a family of subsets of vertices, each referred to as an edge. A set system is merely another representation of a hyper-graph. Hence, the same optimization problems may be phrased either in hyper-graph notation or in set-system notations. Specifically, in the problem of Set-Packing (SP), one is given a family of sets and the problem is to find a maximal sub-family of sets that are pairwise disjoint. This problem is equivalent to the problem of finding a Maximal-Matching in a hyper-graph.

We study this problem in k-uniform hyper-graphs and Independent-Set in graphs of bounded degree.

The bounded degree variants of this problem are known to admit approximation algorithms better than the general versions, the quality of the approximation being a function of the bound on the size of the edges or the bound on the degree of the vertices (detailed below). An extensive body of algorithmic work has been devoted to these restricted problems (for example, [HS89]), but matching inapproximability results have only recently been explored (notably by Trevisan [Tre01]). We continue and explore the problem of k-Dimensional Matching. On our route, we improve existing inapproximability results for k-DM, and thereby improve known inapproximability results for variants of the Independent-Set problem and the Set-Packing problem.

For large k, we are usually interested in the asymptotic dependence of the inapproximability factor on k. However, for small k values, the constant is of major interest. Previous papers typically focus on one of the cases.

We present two main results, one asymptotical inapproximability result (where the hardness factor is a function of the bound), and one for low values of the bound.

1.1 k-Dimensional Matching

The Unbounded Variant. The general problem of finding a matching in a hyper-graph was extensively studied (for example [BYM84, BF94, BF95, BH92, Hås99, Wig83]). Quite tight approximation algorithms and inapproximability results are known for this problem. Håstad [Hås99] proved that Set-Packing cannot be approximated to within $O(N^{1-\varepsilon})$ unless $NP \subseteq ZPP$ (where N is the number of sets). The best approximation algorithm achieves an approximation ratio of $O(\frac{N}{\log^2 N})$ [BH92].

In contrast, the case of bounded variants of this problem seems to be of a different nature.

Bounded Variants. The problem of finding a maximal matching in a bipartite graph *(2-DM)* is known to be solvable in polynomial time, say by a reduction to network flow problems [Pap94]. Polynomial time algorithms are also known for finding a maximal matching in a (not necessarily bipartite) graph [Edm65]. In contrast to the 2-DM case, for all $k \geq 3$ the k-DM problem is NP-hard [Kar72, Pap94]. Furthermore, for $k = 3$, the problem is known to be APX-hard [Kan91], but no explicit approximation hardness factor for 3-DM is known to date. [1]

Currently, the best polynomial time approximation algorithm for k-Set-Packing *(k-SP)* achieves an approximation ratio of $\frac{k}{2}$ (see the result of Hurkens and Schrijver [HS89]). This is, to date, the best approximation algorithm for k-DM as well.

[1]Recently [CC02, BK03] obtained a hardness factor of $\frac{98}{97}$ for 3-DM

Several inapproximability factors were obtained for k-DM, with large values of k. Alon et al [AFWD95] proved that Maximum Independent-Set on graphs with degree bounded by k (denoted k-*IS*) is NP-hard to approximate to within $k^c - \varepsilon$ for some $c > 0$. This result implies the same asymptotical hardness for k-DM by a simple gap preserving reduction (see Appendix E, proposition 45). This was recently improved to the currently best asymptotical inapproximability result by Trevisan [Tre01], who proved that k-IS cannot be approximated to within $\frac{k}{2^{O(\sqrt{\ln k})}}$ unless $P = NP$.
We further improve the inapproximability factor of k-DM, and show that k-DM cannot be approximated to within $O(\frac{k}{\ln k})$ unless $P = NP$. In addition, we improve the known inapproximability results for low k values. Our results, however, do not apply to the problem of k-IS.

1.2 Other variants of Independent-Set

An independent-set in a graph is a matching in its dual hyper-graph. Hence, any hardness factor proven for k-IS, immediately apply to k-DM (but not vice-versa). The best known approximation algorithm for k-IS achieves an approximation ratio of $O(k \log\log k/ \log k)$ (by Vishwanathan [Vis96]). For low bound values instances of k-IS, the best approximation algorithm achieves an approximation ratio of $(k + 3)/5$ for $k \geq 3$ (see [BF94, BF95]). For a survey of approximation algorithms for Independent-Set variants (including bounded degree variant) see [Hal98].
The known asymptotical inapproximability results for k-DM were originally proven for Independent-Set on graphs with degree bound of k. However, unlike k-DM, several explicit inapproximability factors were proven for k-IS with small k values, see [BK99, BK03].
Although hardness factors for k-DM do not immediately apply to k-IS, they do apply to some of its variants.A variant of Independent-Set of interest to us is the Independent-Set in k-claw-free graphs (denoted k-ISCFG). We shall define k-ISCFG formally later, and show that our hardness of approximation factors for k-DM holds for $(k + 1)$-ISCFG.

1.3 Our Results

We prove the following theorems regarding the hardness of k-DM:

Theorem 1 (Asymptotic Hardness) *It is NP-hard to approximate k-DM to within* $O\left(\frac{k}{\ln k}\right)$

Theorem 2 (Hardness for Low Bound Values) *For every* $\varepsilon > 0$ *it is NP-hard to approximate 4-DM, 5-DM and 6-DM to within* $\frac{54}{53} - \varepsilon$, $\frac{30}{29} - \varepsilon$ *and* $\frac{23}{22} - \varepsilon$ *respectively.*

The above two theorems immediately extend to hardness of approximation to within the same factors for the problems of $(k + 1)$-ISCFG [Hal98] and k-SP.
The table below summarizes our inapproximability factors for k-DM and related problems versus known approximation ratios and previous inapproximability results.

1.4 Paper Organization

Section 2 contains the proof of the asymptotic hardness of approximation for k-DM. Section 3 presents the low bound values hardness of approximation for k-DM. Section 4 contains a discussion on the implications of our results, the techniques used and possible directions to improve them. Definitions and restatement of known theorems used throughout this paper can be found in subsections 1.5-1.7 and in appendix B. The proofs regarding hyper-graph disperser and generalized disperser can be found in appendix C and D respectively. Appendix E contains a description of the reduction from k-IS to k-DM.

Problem	Approximation Ratio	Prev. Inapproximability	Our Inapproximability
k-DM $k+1$-ISCFG k-SP	$\frac{k}{2}$ [HS89]	$\frac{k}{2^{O(\sqrt{\ln k})}}$ [Tre01]	$O\left(\frac{k}{\ln k}\right)$
4-DM, 4-SP	2 [HS89]	$\frac{98}{97}-\varepsilon$ [BK03]	$\frac{54}{53}-\varepsilon$
5-DM, 5-SP	$\frac{5}{2}$ [HS89]	$\frac{50}{49}-\varepsilon$ [BK03]	$\frac{30}{29}-\varepsilon$

Table 1: Approximation ratios versus inapproximability results for k-DM and related problems

1.5 Formulation of the Problems Studied

The most general problem of the type studied here, is the Set-Packing problem.

Definition 3 (SP) ***Set-Packing*** *is the following optimization problem:*
Input: *A hyper-graph $H=(V,E)$.*
Problem: *Find a matching of maximal size in H.*

Following are variants of this problem, where some bounds are put on the degree of the vertices and on the size of the edges.

Definition 4 (k-SP) ***k-Set-Packing*** *is the following optimization problem:*
Input: *A k-uniform hyper-graph $H=(V,E)$.*
Problem: *Find a matching of maximal size in H.*

A further restriction, and the main problem studied herein, is the Maximum k-Dimensional Matching (k-DM) problem:

Definition 5 (k-DM) ***k-Dimensional Matching*** *is the following optimization problem:*
Input: *A k-uniform k-partite balanced hyper-graph $H=(V^1,...,V^k,E)$.*
Problem: *Find a matching of maximal size in H .*

Note that a matching in a hyper-graph H is an Independent-Set in its dual graph and vice versa. Hence the following problem is strongly related to k-DM.

Definition 6 (k-IS) ***k-Maximum Independent-Set*** *is the following optimization problem:*
Input: *A graph $G=(V,E)$ with maximal vertex degree bounded by k.*
Problem: *Find an Independent-Set of maximal size in H.*

Due to the previous observation, k-IS can be reduced to $k+1$-DM in a gap-preserving reduction. See Appendix E proposition 45 for details.
The next problem is a generalization of k-IS.

Definition 7 (k-ISCFG) ***Maximum Independent-Set in k Claw Free Graph*** *is the following problem:*
Input: *A graph $G=(V,E)$ which is k claw-free.*
Problem: *Find an Independent-Set of maximal size in H.*

Clearly k-IS is a special case of $(k+1)$-ISCFG, as a graph with maximal degree bound of k is in particular a $(k+1)$-claw-free graph. In addition, the problem of k-DM translates to $k+1$-ISCFG in the dual graph (see [Hal98]).

1.6 Hardness of Approximation

Definition 8 (Gap problems) *Let A be a maximization problem.* ***gap-A-***$[a, b]$ *is the following decision problem:*
Problem: *Given an input instance, decide whether there exists a solution of fractional size at least b, or whether every solution of the given instance is of size smaller than a. If the size of the solution resides between these values, then any output suffices.*

Clearly, for any maximization problem, if gap-A-$[a, b]$ is NP-hard, than it is NP-hard to approximate A to within any factor smaller than $\frac{b}{a}$.
Our main result in this paper is derived by a reduction from the following problem.

Definition 9 (Linear Equations) ***MAX-3-LIN-q*** *is the following optimization problem:*
Input: *A set Φ of linear equations over $GF(q)$, each depending on 3 variables.*
Problem: *Find an assignment that satisfies the maximum number of equations.*

The following central theorem stemmed from a long line of research that formulated in the celebrated PCP theorem (see [ALM+92, AS92]):

Theorem 10 (Håstad [Hås97]) *gap-MAX-3-LIN-q-$[\frac{1}{q} + \varepsilon\ , 1 - \varepsilon\ ]$ is NP-Hard. Furthermore, the result holds for instances of MAX-3-LIN-q in which the number of occurrences of each variable is a constant, chosen from two possible values, and in which no variable appears more then once in a single equation.*

We denote an instance of MAX-3-LIN-q by $\Phi = \{\varphi_1, ..., \varphi_n\}$. Φ is over the set of variables $X = \{x_1, ..., x_m\}$. Each equation has q^2 satisfying assignments. Let $\Phi(x)$ be the set of all equations in Φ depending on x, and $\Phi(x, l)$ be the subset of $\Phi(x)$ where x is the l'th variable in the equation (clearly $l \in [3]$). Note that w.l.o.g. for every $x \in X, \Phi(x, 1) = \Phi(x, 2) = \Phi(x, 3)$ (as we can take three copies of each equation, and shift the location of the variables). Denote by $Sat(\Phi, A)$ the set of all equation of Φ satisfied by an assignment A. If A is an assignment to an equation $\varphi \in \Phi(x)$, we denote by $A[\varphi \to x]$ the corresponding assignment to x.

1.7 Hyper Dispersers

The following definition is a generalization of disperser graphs. For definitions and results regarding dispersers see [RTS00].

Definition 11 *Let $H = (V^1, ..., V^k, E)$ be a k-uniform k-partite balanced hyper-graph, and let $V = \bigcup_{i \in [k]} V^i$. H is called δ-hyper-disperser if for every Independent-Set I, all but δV of the vertices of I are located in one part. Formally, for every Independent-Set $I \subseteq V$ the following holds:*

$$\exists i. I \setminus V^i \leq \delta V$$

2 Proof of the Asymptotic Inapproximability Result for k-DM

This section proves the asymptotic hardness of approximating k-DM (theorem 1) by a deterministic polynomial time reduction from MAX-3-LIN-q to k-DM. In the following construction we utilize the existence of disperser-like graphs, as stated in the following lemma.

Definition 12 (Hyper-Graph Edge-Disperser) *Let $D(t,q)$ be a hyper-graph with the following properties; $D(t,q)$ is q-regular d-uniform (where $d = \lceil 3q \ln q + 2q \rceil$). It has dt vertices and qt edges; It is d-partite balanced (each part of size t); it is q-strongly-edge-colorable; and it is a dual graph to a $\frac{1}{q^2}$-hyper-disperser.*

Lemma 13 *For every $q > 1$ and $t > 1$ there exists a hyper-graph $D(t,q)$ with the above parameters.*

For proof see appendix C.

2.1 The construction

Let $\Phi = \{\varphi_1, ..., \varphi_n\}$ be an instance of MAX-3-LIN-q over the sets of variables X and Y. Theorem 10 holds even if each variable $x \in X$ and $y \in Y$ occurs a constant number of times c_X and c_Y respectively. Furthermore, the number of a variable's occurrences in the first second and third location of an equation is equal. We now describe how to deterministically construct, in polynomial time, an instance of k-DM - the hyper-graph $H_\Phi = (V, E)$.

Let $D_X \equiv D(c_X, q)$ and $D_Y \equiv D(c_Y, q)$ (as stated in lemma 13). For every variable $x \in X$ (or $y \in Y$) we have a copy D_x of D_X (or D_y of D_Y). Formally, for every $x \in X \cup Y$,

$$V(D_x) = \{v_{x,\varphi,i} | \varphi \in \Phi(x), i \in [d]\}$$

$$E(D_x) = \{e_{x,\varphi,i} | \varphi \in \Phi(x), i \in [q]\}$$

where the index $i \in [q]$ is given by a coloring of the edges with q colors such that no two edges of the same color share a vertex (recall that such a coloring exists as this graph is q-strongly-edge-colorable).

The Vertices of H_Φ. The vertices of H_Φ are the union of the vertices of all the copies of D_X and D_Y namely,

$$V = X \times V(D_X) \bigcup Y \times V(D_Y)$$

that is

$$V = \{v_{x,\varphi,i} | x \in X \cup Y, \varphi \in \Phi(x), i \in [q]\}$$

The Edges of H_Φ. We have an edge for each equation $\varphi \in \Phi$ and a satisfying assignment to it. Consider an equation $\varphi = x_1 + x_2 + x_3 = a \mod q$, and a satisfying assignment A to that equation (note that there are q^2 such assignments, as assigning the first two variables, determines the third). The corresponding edge, $e_{\varphi,A}$, is composed of three edges, one from the hyper-graph D_{x_1}, one from D_{x_2} and the last from D_{x_3}. Formally:

$$E = \{e_{\varphi,A} | A \in [q]^3, \varphi \in Sat(\Phi, A)\}$$

$$e_{\varphi,A} = e_{x,\varphi,A(x)} \cup e_{y,\varphi,A(y)} \cup e_{z,\varphi,A(z)}$$

Where $A(x_1), A(x_2), A(x_3)$ are the restrictions of the assignment A to the variables x_1, x_2, x_3 respectively (note that each of the three composing edge, participates in creating q edges). Clearly, the cardinality of $e_{\varphi,A}$ is $3d$.
This concludes the construction.

Notice that the construction is indeed deterministic, as each variable occurs a constant number of times (see theorem 10). Hence, the size of D_X and D_Y is constant and its existence (see lemma 13) suffices, as one can enumerate all possible hyper-graphs, and verify their properties.

Proposition 14 *H_Φ is $3d$-partite-balanced.*

Proof: We show how to partition V into $3d$ independent sets of equal size. Let the sets be $P_{l,i}$ whereas $i \in [d]$ and $l \in [3]$:

$$P_{l,i} = \{v_{x,\varphi,i} | x \in X \cup Y, \varphi \in \Phi(x,l)\}$$

$P_{l,i}$ is clearly a partition of the vertices, as each vertex belongs to a single part.
We now explain why each part is an independent set. Let $P_{l,i}$ be an arbitrary part, and let $e_{\varphi,A} \in E$ be an arbitrary edge, where $\varphi \equiv x_1 + x_2 + x_3 = a \mod q$:

$$e_{\varphi,A} = e_{x,\varphi,A(x_1)} \cup e_{y,\varphi,A(x_2)} \cup e_{z,\varphi,A(x_3)}$$

$P_{l,i} \cap e_{\varphi,A}$ may contain vertices corresponding only to one of the variables x_1, x_2, x_3, since it contains variables corresponding to a single location (first, second or third).
Let that variable be, w.l.o.g, x_1. The edge $e_{x_1,\varphi,A(x_1)}$ contains exactly one vertex from each of the d parts, as the graph D_{x_1} is d-partite. Therefore, the set $P_{l,i} \cap e_{\varphi,A}$ contains exactly one vertex.
Since $|P_{l,i} \cap e_{\varphi,A}| = 1$ for every edge and every set $P_{l,i}$, the graph H_Φ is $3d$-partite-balanced. ■

Claim 15 *[Completeness] If there is an assignment to Φ which satisfies $1-\varepsilon$ of its equations, then there is a matching in H_Φ of size $\left(\frac{1-\varepsilon}{q^2}\right) E$.*

Proof: Let A be an assignment that satisfies $1-\varepsilon$ of the equations. Consider the matching $M \subseteq E$ consisted of all edges corresponding to A, namely

$$M = \{e_{\varphi,A(\varphi)} | \varphi \in Sat(\Phi, A)\}$$

Trivially, $M = \left(\frac{1-\varepsilon}{q^2}\right) E$, as we took one edge corresponding to each satisfied equation. These edges are indeed a matching since for each variable, only edges corresponding to a single assignment to that variable are taken. ■

Lemma 16 *[Soundness] If every assignment to Φ satisfies at most $\frac{1}{q}+\varepsilon$ fraction of its equations, then every matching in G is of size $O\left(\frac{1}{q^3}E\right)$.*

Proof: Denote by E_x the edges of H_φ corresponding to equations φ containing the variable x, namely,

$$E_x = \{e_{\varphi,A} | \varphi \in \Phi(x), A \in [q^2]\}$$

Denote by $E_{x=a}$ the subset of E_x corresponding to an assignment of a to x, that is,

$$E_{x=a} = \{e_{\varphi,A} | \varphi \in \Phi(x), A[\varphi \to x] = a\}$$

Let M be a matching of H_Φ.
Let A_{maj} be the most popular assignment. That is, for every $x \in X \cup Y$ choosing the assignment of x to be such that it corresponds to maximal number of edges. Formally, choose

$$A_{maj}(x) \in [q] \text{ s.t. } |E_{x=a} \cap M| \text{ is maximized}$$

Let M_{maj} be the set of edges in M that agree with A_{maj}, and M_{min} be all the other edges in M, namely

$$M_{maj} = \{e_{\varphi,A_{maj}}\}_{\varphi\in\Phi}$$

$$M_{min} = M \setminus M_{maj}$$

As $Sat(\Phi, A_{maj}) \leq \frac{1}{q} + \varepsilon$, we have $M_{maj} < (\frac{1}{q} + \varepsilon)\frac{E}{q^2}$.
From the disperser-properties of D_X and D_Y (derived from lemma 13) we know that

$$M_{min} \cap E_x \leq \frac{1}{q^2} E(D_x)$$

This means that

$$M_{min} \cap E_x \leq \frac{1}{q^3} E_x$$

as every edge of D_x is a subset of q hyper edges in E_x, but only one of such q edges can be taken to M as they share vertices (recall that M is a matching).
Therefore,

$$M_{min} \leq \sum_{x\in X} M_{min} \cap E_x \leq \frac{1}{q^3} \sum_{x\in X} E_x = \frac{1}{q^3} E$$

and thus

$$M = M_{min} + M_{maj} \leq \frac{2}{q^3} E$$

■

By claim 15 and lemma 16 we showed that Gap-k-DM-$\left[\frac{2}{q^3}, \frac{1}{q^2} - \varepsilon\right]$ is NP-hard. Since each edge is of size $k = 3d = 9\lceil q \ln q + 2q\rceil$ it is NP-hard to approximate k-DM to within $O(\frac{\ln k}{k})$.

3 Proof of the Low k Values Inapproximability result for k-DM

3.1 The construction

We show a deterministic polynomial time reduction from gap-MAX-3-LIN-2-$\left[\frac{1}{2}+\varepsilon, 1-\varepsilon\right]$ to gap-k-DM-$[f(k)+\varepsilon, 1-\varepsilon]$, where $f(k)$ is a constant that satisfies: $\forall k \geq 4$. $\frac{7}{8} < f(k) < 1$, and specifically:

$$f(4) < \frac{53}{54}\ ,\ f(5) < \frac{29}{30}, f(6) < \frac{22}{23}$$

The formal description of the reduction is somewhat tedious and perhaps misleading of its simplicity. Therefore, we first give an overview of the construction, and then provide the formal definitions and propositions.

Overview

For $\Phi = \{\varphi_1, ..., \varphi_n\}$, an instance of Lin-3-eq-2 over the set of variables $X = \{x_1, ..., x_m\}$, we construct a k-uniform k-partite balanced hyper-graph $H_\Phi = (V^1, ..., V^k, E)$.
The construction is composed of a *consistency gadget* and a *gap gadget*.
The first $k-1$ parts of H_Φ are the consistency gadget. These are further partitioned into sets corresponding to the different variables. For each variable, we construct two sets of edges that relate to its two possible assignments. The construction implies that a matching that contains a large set edges of one type can contain only small set of edges of the other type. Thus, this gadget enforces the consistency notion.
The gap gadget is the scheme in which the edges contain vertices of the last part. This scheme enables a large matching in case that the equation set has an assignment that satisfies almost all of its equations. In case that any assignment to the equation set satisfies only a small fraction of its equations, the gap gadget permits only a small matching.
For detailed construction and proofs, see appendix A.

4 Discussion

Asymptotic k-DM

The gap between the asymptotic inapproximability result presented herein for k-DM, and the tightest approximation algorithm known, was reduced to $O(\log k)$. The question whether the best approximation ratio achievable in polynomial time is $\frac{k}{2}$ or $O(\frac{k}{\log k})$ is interesting by itself, as well as its implications to the difference between k-DM and k-IS.
The current asymptotic inapproximability factor of $O(\frac{k}{\log k})$ for k-DM approaches the tightest approximation ratio known for k-IS, namely $O(k \log\log k / \log k)$ [Vis96]. Thus, a small improvement in either the approximation ratio or the inapproximability factor will show these problems to be of inherently different hardness.
An improvement in the low bound values hardness factor for k-DM may also separate these problems. The tightest known approximation algorithm for low bound values of k-IS achieves an approximation ratio of $(k+3)/5$ for $k \geq 3$ [BF94, BF95]. Thus, improving the low bound values result up to factors of $\frac{6}{5}+\varepsilon$ for 3-DM or $\frac{7}{5}+\varepsilon$ for 4-DM, suffices to separate these problems.

Properties and Limitations of the Technique

An interesting property of the asymptotic result is regularity of the hyper-graph. That is, the hyper-graph H_Φ is $\Theta\left(\frac{k}{\ln k}\right)$-regular (recall H_Φ was q^2-regular where the parameter q refers to the size of the field of the Max-3-Lin-q instance). As for the low bound values result, the hyper-graph H_Φ is not regular, but its vertices degrees are distributed evenly: half of the vertices have degree 2 and half of degree 4.

Another interesting property of the construction (for both asymptotic and low bound values results) is called *almost perfect completeness*. This property refers to the fact that the matching proved to exist in the completeness claims 15 and 18 is almost a perfect matching. Knowing the location of a gap is interesting by itself and may proof useful (in particular if it is extreme on either the completeness or the soundness). On the significance of this property refer to [Pet94].

A limitation of the result is the lack of inapproximability factor for 3-DM. This is, in fact, an inherent limitation of the technique. The proof of soundness is composed of two main sections: limiting the size of the majority fraction of edges in a matching, and using the minority fraction in a set of edges to derive a large probability for collision. For the first part we use the gap gadget, and for the second part we analyze the consistency gadget. A well known fact is that any regular extractor (and thus a disperser which is special case of an extractor) must have vertices degrees of at least 3. Thus, our consistency gadget, which is a generalization of a disperser (see appendix D), has at least 3 columns. As the gap gadget is separated from the consistency gadget, this forces a minimum of $k \geq 4$ columns in the construction.

Improving the Low Bound Values Result

Obtaining tighter inapproximability results for small k values may be more difficult. This is the case for low bound values instances of other problems (for example, for the Vertex Cover problem, a long line of complex proofs and ideas let to the current state of inapproximability factor slightly larger then $\frac{4}{3}$ [DS02] versus an approximation ratio of 2).

The current gap between known approximation ratios and unapproximability factors for k-DM is significant. For $k = 4$ the current gap is 2 vs. $\frac{54}{53}$, and for $k = 3$ the known approximation ratio is $\frac{3}{2}$ but no explicit inapproximability factor is known to date[1].

A plausible direction to improve the inapproximability results for k-DM is to tighten the analysis of the soundness lemma. In the soundness analysis, the calculations of the maximal size of a matching in the graph H_Φ completely disregards the last column of vertices. The proof takes into account only the first $k-1$ columns of the consistency gadget. However, the last column also introduces a restriction on the matchings in the graph. This additional restriction supplied by the last column, which is not utilized hereby, may strengthen our result and perhaps supply an inapproximability result for 3-DM. This is impossible with our current technique of analyzing only the consistency gadget of $k-1$ columns as explained above.

A second direction is to replace the gap gadget. A gadget is denoted an α-gadget if there is a factor of α between the number of vertices that can be covered by a matching that corresponds to a satisfying assignment, and the number of vertices that can be covered by any set of edges that correspond to an unsatisfying assignment. The current gadget is composed of four vertices. All four vertices can be covered by a matching that corresponds to a satisfying assignment, but at most three out of four can be covered by any set of edges that correspond to an unsatisfying assignment (refer to proposition 17). Thus, our gadget is a $\frac{4}{3}$-gadget. One might attempt to

[1]Recently [CC02, BK03] obtained a hardness factor of $\frac{98}{97}$ for 3-DM

construct a different gadget, applicable to a reduction from the more general problem of Max-3-Lin-q, and obtain a gadget with a larger α parameter. However, any α-gadget applicable a reduction from Max-3-Lin-q is bounded by $\alpha \leq \frac{3}{2}$. The reason is the requirement from the gap gadget that a set of edges corresponding to a satisfying assignment be a matching. Because any variable can determine whether a certain assignment is satisfying or not, a set of edges for any two variables out of the three for a certain equation, must constitute a matching. Thus, for any set of edges for an equation, edges corresponding to any one variable can be removed, and the remaining edges must constitute a matching. Therefore α is bound by $\frac{3}{2}$.

References

[AFWD95] N. Alon, U. Fiege, A. Wigderson, and D.Zuckerman. Derandomized graph products. *Computational Complexity*, 5:60–75, 1995.

[ALM+92] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. In *Proc. 33rd IEEE Symp. on Foundations of Computer Science*, pages 13–22, 1992.

[AS92] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. In *Proc. 33rd IEEE Symp. on Foundations of Computer Science*, pages 2–13, 1992.

[AS00] N. Alon and J. H. Spencer. *The Probabilistic Method*. John Wiley and Sons, New York, 2000.

[BF94] P. Berman and M. Furer. Approximating maximum independent set in bounded degree graphs. *SODA*, pages 365–371, 1994.

[BF95] P. Berman and T. Fujito. On the approximation properties of independent set problem in degree 3 graphs. *WADS*, pages 449–460, 1995.

[BH92] R. Boppana and Magnus M. Halldorsson. Approximating maximum independent sets by excluding subgraphs. *Bit 32*, pages 180–196, 1992.

[BK99] P. Berman and M. Karpinski. On some tighter inapproximability results. *DIMACS Technical Report 99-23*, 1999.

[BK03] P. Berman and M. Karpinski. Improved approximation lower bounds on small occurrence optimization. *Manuscript*, 2003.

[BYM84] R. Bar-Yehuda and S. Moran. On approximation problems related to the independent set and vertex cover problems. *Discrete Applied Mathematics*, 9:1–10, 1984.

[CC02] M. Chlebik and J. Chlebikova. Approximation hardness for small occurrence instances of np-hard problems. *ECCC TR02-073*, 2002.

[DS02] I. Dinur and S. Safra. The importance of being biased. *In Proc. of the 34rd ACM STOC*, 2002.

[Edm65] J. Edmonds. Paths, trees and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.

[Hal98] Magnus M. Halldorsson. Approximations of independent sets in graphs. *APPROX*, 1998.

[Hås97] Johan Håstad. Some optimal inapproximability results. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 1–10, El Paso, Texas, 4–6 May 1997.

[Hås99] Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Math.*, 182(1):105–142, 1999.

[HS89] C. A. J. Hurkens and A. Schrijver. On the size of systems of sets every t of which have an sdr, with an application to the worst-case ratio of heuristics for packing problems. *SIAM Journal Discrete Math*, 2:68–72, 1989.

[Kan91] V. Kann. Maximum bounded 3-dimensional matching is maxsnp-complete. *Information Processing Letters*, 37:27–35, 1991.

[Kar72] R. M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, pages 83–103, 1972.

[Pap94] C. Papadimitriou. *Computational Complexity*. Addison Wesley, 1994.

[Pet94] E. Petrank. The hardness of approximation - gap location. *Israel Symposium on Theory of Computing Systems*, 1994.

[RTS00] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13:2–24, 2000.

[Tre01] L. Trevisan. Non-approximability results for optimization problems on bounded degree instances. *In Proc. of the 33rd ACM STOC*, 2001.

[Vis96] S. Vishwanathan. Personal communication to m. halldorsson cited in [Hal98]. 1996.

[Wig83] A. Wigderson. Improving the performance guarantee for approximate graph coloring. *Journal of the Association for Computing Machinery*, 30(4):729–735, 1983.

A Proof of the Low k Values Inapproximability result for k-DM

A.1 The construction

We begin by describing the construction of a $(k-1)$-uniform $(k-1)$-partite balanced hyper-graph $\hat{D}_\Phi = (V^1, ..., V^{k-1}, F)$. This graph is the consistency gadget. We later explain how to extend this graph to $H_\Phi = (V^1, ..., V^k, E)$ by applying the gap gadget.

The consistency gadget: The hyper-graph $\hat{D}_\Phi$

The Vertices of $\hat{D}_\Phi$. Denote the vertices of the hyper-graph $\hat{D}_\Phi$ by:

$$V^i = \{v_x^i(l) | x \in X, l \in [d_x]\}$$

$$V(\hat{D}_\Phi) = \bigcup_{i \in [k-1]} V^i$$

Where $d_x = \frac{4}{3} \cdot \Phi(x)$.
Denote by V_x the set of vertices that correspond to a certain variable $x \in X$:

$$V_x = \{v_x^i(l) \in \cup_j V^j\}$$

Each set V_x is partitioned to $k-1$ different parts, $V_x^1, ..., V_x^{k-1}$:

$$\forall i \in [k-1] \ . \ V_x^i = V^i \cap V^x$$

The Edges of $\hat{D}_\Phi$. In this paragraph we only name the edges. In the next paragraph we specify which vertices are contained in each edge.
For every set of vertices V_x we construct two sets of edges, $F_x = F_{x=0} \cup F_{x=1}$, each corresponding to a possible assignment to x. Namely:

$$F_{x=1} = \{e_{x=1}^i | i \in [d_x]\}$$

$$F_{x=0} = \{e_{x=0}^i | i \in [d_x]\}$$

$$F = \bigcup_{a \in \{0,1\}, x \in X} F_{x=a}$$

Altogether that's $2m$ sets of edges.

Structure of $\hat{D}_\Phi$. According to lemma 31, for every $\varepsilon > 0$ there exists $(\Omega_{k-1} + \varepsilon)$-dispersers that are (k-1)-regular and (k-1)-strongly-edge-colorable. Define the graph $D_x = (F_{x=0}, F_{x=1}, V_x)$ to be such an $(\Omega_{k-1} + \varepsilon)$-disperser on d_x vertices (recall that the number of vertices of this disperser must satisfy: $\geq d_x \cdot T$ vertices for some constant T. We can assume that w.l.o.g. $T = 1$, as otherwise we could duplicate T times the equation set Φ).
Let $\hat{D}_x = (V_x^1, ..., V_x^{k-1}, F_x)$ be the dual graph of D_x. Since D_x is (k-1)-regular and (k-1)-strongly-edge-colorable, its dual, $\hat{D}_x$, is (k-1)-uniform (k-1)-partite balanced.
Notice that this construction is indeed deterministic, as each $x \in X$ occurs a constant number of times (see theorem 10). Hence, the size of each $(\Omega_{k-1} + \varepsilon)$-disperser is constant, and one can find such graphs by enumerating all possible graphs.

Introducing the gap gadget: constructing H_Φ

In this subsection we explain how to transform $\hat{D}_\Phi$ into H_Φ by using the gap gadget.

The vertices V^k. The graph $H_\Phi = (V^1, ..., V^k, E)$ contains the same vertices as $\hat{D}_\Phi$, and, in addition, the vertices V^k that constitute a new part. Denote these vertices as:

$$V^k = \{v^k_\varphi[b] | \varphi \in \Phi, b \in [4]\}$$

The vertices that correspond to a certain equation $\varphi \in \Phi$ are denoted by V^k_φ where:

$$V^k_\varphi = \{v^k_\varphi[b] | b \in [4]\}$$

Notice that the number of vertices in V^k is:

$$V^k = \sum_{\varphi \in \Phi} V^k_\varphi = 4n = \frac{4}{3} \cdot 3m$$

The connection scheme The vertices V^k are added to the edges $F = E(\hat{D}_\Phi)$ so to create the k-uniform edges $E = E(H_\Phi)$. The connection scheme is identical for each equation $\varphi \in \Phi$. For each equation $\varphi = x \oplus y \oplus z = t$, $t \in \{0, 1\}$, we create a set of twelve edges denoted E_φ by amending eight edges from F:

1. Arbitrarily pick two edges from $F_{x=0}$ and add the vertices $v^k_\varphi[1], v^k_\varphi[2]$ to them, one vertex to each.
2. Arbitrarily pick two edges from $F_{x=1}$ and add the vertices $v^k_\varphi[3], v^k_\varphi[4]$ to them, one vertex to each.
3. Arbitrarily pick one edge from $F_{y=0}$, duplicate it, and add the vertices $v^k_\varphi[1], v^k_\varphi[3]$, one to each copy (thus creating two edges).
4. Arbitrarily pick one edge from $F_{y=1}$, duplicate it, and add the vertices $v^k_\varphi[2], v^k_\varphi[4]$, one to each copy (thus creating two edges).
5. Arbitrarily pick one edge from $F_{z=0}$, duplicate it, and if $t = 0$ add the vertices $v^k_\varphi[1], v^k_\varphi[4]$, one to each copy (thus creating two edges). If $t = 1$ then add the vertices $v^k_\varphi[2], v^k_\varphi[3]$ instead.
6. Arbitrarily pick one edge from $F_{z=1}$, duplicate it, and if $t = 0$ add the vertices $v^k_\varphi[2], v^k_\varphi[3]$, one to each copy (thus creating two edges). If $t = 1$ then add the vertices $v^k_\varphi[1], v^k_\varphi[4]$ instead.

Of course, at each stage we pick edges that were not picked before. Denote by $E_x, E_{x=0}, E_{x=1}$ the edges that were created from the sets of edges $F_x, F_{x=0}, F_{x=1}$ respectively. In addition, denote all twelve edges corresponding to a certain equation as:

$$E_\varphi = \{e \in E | e \cap V^k_\varphi \neq \emptyset\}$$

The connection scheme can be summarized by the following table. The columns stand for the different variables of the equation $\varphi = x \oplus y \oplus z = 0$. The rows represent the possible assignments

to a variable. A $\wedge$ sign implies that two edges are constructed, one for each vertex (see stages 1-2 of the process described above). A $\vee$ sign implies that a single edge is duplicated to contain each vertex (see stages 3-6 of the process).
A similar table describes the connection scheme for equations of the form $\varphi = x \oplus y \oplus z = 1$. For the latter type of equations, the two cells of the last column are swapped.

Assignment / Variable	x	y	z
0	$v_\varphi^k[1] \wedge v_\varphi^k[2]$	$v_\varphi^k[1] \vee v_\varphi^k[3]$	$v_\varphi^k[1] \vee v_\varphi^k[4]$
1	$v_\varphi^k[3] \wedge v_\varphi^k[4]$	$v_\varphi^k[2] \vee v_\varphi^k[4]$	$v_\varphi^k[2] \vee v_\varphi^k[3]$

Table 2: The connection scheme for an equation $\varphi = x \oplus y \oplus z = 0$

For any equation $\varphi \in \Phi$ and any assignment to the variables $A : X \mapsto \{0,1\}$, denote the six edges from E_φ corresponding to an assignment A as $E_{A,\varphi}$.

Proposition 17 *For every $\varphi \in \Phi$ and an assignment $A : X \mapsto \{0,1\}$, it holds that $\varphi \in Sat(\Phi, A)$ if and only if there exists a matching $M_{A,\varphi} \subseteq E_{A,\varphi}$ of size $M_{A,\varphi} = 4$.*

Proof: Let $\varphi \equiv x \oplus y \oplus z = t$. Notice that the boolean expression in table 2 that corresponds to A contains four different vertices if and only if the assignment is satisfying. Hence, there is no matching of cardinality four for an unsatisfying assignment.
Furthermore, notice that if the assignment is satisfying, then from the six edges of $E_{A,\varphi}$, the two edges corresponding to x, one edge corresponding to y, and one to z cover all of V_φ^k. According to the construction, these four edges are a matching. ■

A.2 Definitions

Consider a matching $M \subset E$.
Let $A_{maj} : X \mapsto \{0,1\}$ be the majority assignment, namely:

$$A_{maj}(x) := \begin{cases} 1 & M \cap E_{x=1} \geq M \cap E_{x=0} \\ \\ 0 & otherwise \end{cases}$$

Similarly, define as $A_{min} : X \mapsto \{0,1\}$:

$$A_{min}(x) := 1 - A_{maj}(x) \text{ (the complementary assignment).}$$

We partition the matching M into two sets. Define M_{maj} to be the set of all edges of M corresponding to the majority assignment A_{maj}. Let M_{min} be the all other edges of M. Namely:

$$M_{maj} := M \cap E_{A_{maj}}$$

$$M_{min} := M \setminus M_{maj} = M \cap E_{|A_{min}}$$

Let β be the fractional size of the set M defined as:

$$\beta = \frac{k \cdot M}{V}$$

Notice that β is defined to be a fraction of the vertices, as the graph H_Φ is k-uniform.
Let β_{maj} be the fractional size of the majority set from the matching - M_{maj}, and β_{min} be the fractional size of M_{min}:

$$\beta_{maj} = \frac{kM_{maj}}{V}$$

$$\beta_{min} = \beta - \beta_{maj} = \frac{kM_{min}}{V}$$

Similarly, for a certain variable x, let β_x be the fractional size of the set $M \cap E_x$, and $\beta_{x,min}, \beta_{x,maj}$ the fractional sizes of the minority and majority defined as:

$$\beta_x = k\tfrac{M \cap E_x}{V_x}$$
$$\beta_{x,min} = k\tfrac{M_{min} \cap E_x}{V_x}$$
$$\beta_{x,maj} = k\tfrac{M_{maj} \cap E_x}{V_x}$$

A.3 Completeness

Claim 18 (Completeness) *If there is an assignment to Φ that satisfies $1-\varepsilon$ fraction of its equations, then there exists a matching $M \subseteq E$ of size $M \geq \frac{1-\varepsilon}{k}V$.*

Proof: Let $A : X \mapsto \{0,1\}$ be an assignment to Φ which satisfies $1-\varepsilon$ fraction of its equations. Define the set M to be the union of all matchings $M_{A,\varphi}$ corresponding to the equations satisfied by A. Formally:

$$M = \bigcup_{\varphi \in Sat(A,\Phi)} M_{A,\varphi}$$

Obviously, the edges of M do not intersect on the k'th part of H_Φ, since two edges may collide on this column only if they correspond to the same equation, but for each equation a matching was taken. Hence, it remains to show that no two edges of M share a vertex from the first $k-1$ parts of H_Φ.
Denote by $M|_F$ the set of edges of $\hat{D}_\Phi$ defined by the edges of M removed of their vertex from the k'th part. It remains to show that $M|_F$ is a matching.
Notice that two edges may share vertices from the first $k-1$ columns only if they correspond to the same variable. Consider two edges e_1, e_2 corresponding to the same variable x. Since they both correspond to the majority assignment, the vertices of the dual graph D_x corresponding to these edges reside in the same part. Thus these two edges are either disjoint or equivalent on the first $k-1$ columns. Since the matching M contains no duplicated edges, e_1 and e_2 are pairwise disjoint. Hence M is a matching.
Observe the last column. For every equation φ that is satisfied by A, all four vertices V^φ are covered by M (see proposition 17). Thus, at least $1-\varepsilon$ of all vertices in the last column are covered by M. Since M is a matching and H_Φ is k-uniform k-partite balanced (namely each edge contains exactly one representative from each column), the fraction of vertices M covers on every column is equal. Hence, M covers at least $1-\varepsilon$ of the vertices of H_Φ. ∎

A.4 Soundness

Lemma 19 (Soundness) *If every assignment to Φ satisfies at most a fraction $\frac{1}{2}+\varepsilon$ of its equations, the following holds: Every matching $M \subseteq E$ is of size at most $M \leq (f(k)+\varepsilon)\frac{V}{k}$,* *where $f(k)$ is a constant that for every $k \geq 4$ satisfies: $\frac{7}{8} < f(k) < 1$. Specifically:*

$$f(4) < \frac{53}{54}$$

$$f(5) < \frac{29}{30}$$

$$f(6) < \frac{22}{23}$$

Proof: Let $M \subseteq E$ be a matching in H_Φ. We shall bound the size of its parts M_{maj} and M_{min}. We begin with a bound on the size of the majority:

Claim 20 $M_{maj} < (\frac{7}{8} + \varepsilon) \cdot \frac{V}{k}$

Proof: By proposition 17, for every equation $\varphi \in \Phi$ that is not satisfied by the majority assignment, at most three vertices are covered by M, namely: $M_{maj} \cap M_\varphi \leq 3$.
Since every assignment satisfies at most a fraction $\frac{1}{2} + \varepsilon$ of the equations, the fraction of vertices that the edges of M_{maj} can cover is most:

$$(\frac{1}{2} + \varepsilon) \cdot 1 + (\frac{1}{2} - \varepsilon) \cdot \frac{3}{4} = \frac{7}{8} + \frac{\varepsilon}{4}$$

■

We proceed with a bound on the minority:

Claim 21 *For every $k \geq 4$, $M_{min} < \left(\frac{1-\beta}{1-\alpha_{k-1}} \cdot \frac{\alpha_{k-1}}{2} + \varepsilon \right) \cdot \frac{V}{k}$.*

Proof: For every variable x, the set $M_{x,min}$ is a matching (being a subset of the matching M). Thus, the edges of $M_{x,min}$ removed of their vertices from the k'th column are a matching in the graph $\hat{D}_x$. Therefore, the corresponding vertices to the edges of $M_{x,min}$ constitute an independent set in the dual graph D_x. As D_x is an $(\Omega_{k-1} + \varepsilon)$-disperser, it follows that:

$$\forall_{x \in X} \;.\; \beta_{x,min} < \Omega_{k-1}(\beta_x) + \varepsilon$$

$$M_{min} \quad < \quad \sum_{x \in X} (\Omega_{k-1}(\beta_x) + \varepsilon) \cdot \frac{V_x}{k} \leq$$

W.l.o.g. $M \geq \alpha_{k-1} \cdot \frac{V}{k}$ (otherwise we have a better bound on the size of M then the bound that the lemma proposes). Recall that Ω_{k-1} is smaller then the line $l_{k-1}(x)$ in the range $(\alpha_{k-1}, 1)$ (see definition 30). Hence, according to the convexity of l_{k-1}:

$$\leq \quad \sum_{x \in X_1} (l_{k-1}(\alpha_{k-1}) + \varepsilon) \cdot \frac{V_x}{k} + \sum_{x \in X_2} (l_{k-1}(1) + \varepsilon) \cdot \frac{V_x}{k}$$

Where $X = X_1 \cup X_2$ is a partition of the variables that maximizes the above sum. As $l_{k-1}(1) = \Omega_{k-1}(1) = 0$ and $l_{k-1}(\alpha_k) = \Omega_{k-1}(\alpha_k) = \frac{\alpha_k}{2}$, we get

$$M_{min} < \left(\frac{\alpha_k}{2} + 2\varepsilon \right) \sum_{x \in X_1} \frac{V_x}{k} \tag{1}$$

Where:

$$\sum_{x\in X_1} \alpha_{k-1} \cdot \frac{V_x}{k} + \sum_{x\in X_2} \cdot \frac{V_x}{k} = M$$

Namely:

$$\sum_{x\in X_1} \frac{V_x}{k} = \frac{1-\beta}{1-\alpha_{k-1}} \cdot \frac{V}{k}$$

Hence by equation 1:

$$M_{min} < \left(\frac{1-\beta}{1-\alpha_{k-1}} \cdot \frac{\alpha_{k-1}}{2} + \varepsilon \right) \cdot \frac{V}{k}$$

■

Using claims 20 and 21 we get the following bound on M:

$$\beta \cdot \frac{V}{k} = M = M_{maj} + M_{min} < \left(\frac{7}{8} + \varepsilon \right) \frac{V}{k} + \left(\frac{1-\beta}{1-\alpha_{k-1}} \cdot \frac{\alpha_{k-1}}{2} + \varepsilon \right) \frac{V}{k}$$

Thus:

$$\beta \leq \frac{7 - 3\alpha_{k-1}}{8 - 4\alpha_{k-1}} + \varepsilon \equiv f(k) + \varepsilon$$

Substituting the values for α_k as derived in lemma 31, namely $\alpha_3 \leq 0.919, \alpha_4 \leq 0.842, \alpha_5 \leq 0.778$ we get:

$$f(4) < \frac{53}{54}, f(5) < \frac{29}{30}, f(6) < \frac{22}{23}$$

■

By claim 18 and lemma 19 Gap-k-DM-$[1-\varepsilon, f(k)+\varepsilon]$ is NP-hard. Hence k-DM cannot be approximated to within $O\left(f(k)+\varepsilon\right)$ unless $P = NP$.

B Preliminaries

B.1 Some Basic Notions

For the sake of simplicity, the cardinality of a set S is denoted by S (instead of $|S|$), where ever this meaning is easily understood from the context.

Definition 22 (Hyper-Graph) *A hyper-graph $H = (V, E)$ is a set of vertices V and a family E of subsets of V. Each member of E is called an edge.*

Definition 23 (Matching) *Let $H = (V, E)$ be a hyper-graph. A subset $M \subseteq E$ is called a* matching *if all its edges are pairwise disjoint, namely,*

$$\forall e_1, e_2 \in M \quad e_1 \cap e_2 = \emptyset$$

A matching M is of maximal size, *if for every other matching M', we have*

$$M' \leq M$$

We call a matching M perfect *if it covers all the vertices of a the graph, that is,*

$$\bigcup_{e \in M} e = V$$

Definition 24 (Independent-Set) *Let $H = (V, E)$ be a hyper-graph. A subset of vertices $I \subseteq V$ is called an* Independent-Set *if no two vertices in it are neighbors, that is,*

$$\forall e \in E \quad |e \cap I| \leq 1$$

An Independent-Set I is of maximal size, *if for every other Independent-Set I' we have*

$$I' \leq I$$

We next define regular, uniform, partite, balanced and claw-free hyper-graphs.

Definition 25 (Bounds on Hyper-Graphs) *Let $H = (V, E)$ be a hyper-graph.*
*H is called k-*uniform *if all its edges are of cardinality exactly k, that is,*

$$\forall e \in E . |e| = k$$

*H is called k-*partite *if there is a partition of V into k sets, such that each part is an Independent-Set. H may be denoted:*

$$H = (V^1, V^2, ..., V^k, E)$$

Hence, in a k-uniform k-partite hyper-graph $H = (V^1, V^2 ..., V^k, E)$, we have

$$E \subseteq V^1 \times ... \times V^k$$

A k-partite hyper-graph $H = (V^1, V^2 ..., V^k, E)$ is called balanced *if all its parts are of equal size, that is,*

$$|V^1| = |V^2| = ... = |V^k|$$

Notice that if there exists a perfect matching in a k-uniform k-partite hyper-graph H, then H is balanced.

*A hyper-graph H is called d-*regular *if the degree of each vertex is exactly d.*
*A hyper-graph H is called d-*strongly-edge-colorable *if there exists a coloring of the edges $f : E \mapsto [d]$ so that each vertex participates in at most one edge of each color. Formally:*

$$\forall e_1, e_2 \in E, e_1 \cap e_2 \neq \emptyset . f(e_1) \neq f(e_2)$$

*A k-*claw *is a graph of $k+1$ vertices and k edges. One node is of degree k and all other vertices are of degree 1. A graph $G = (V, E)$ is called k-*claw-free *if there is no induced subgraph which is a k-claw.*

Definition 26 *For a hyper-graph $H_1 = (V, E)$, define its* dual graph *$H_2 = (E, V)$ to be the hyper-graph in which vertices are the edges of H_1 and edges are vertices of H_1. For the graph H_2, every edge $v \in V$ contains all vertices $e \in E$ such that $v \in e$ in H_1.*

Notice that a hyper-graph is k-partite if and only if its dual is k-strongly-edge-colorable and the hyper-graph is k-regular if and only if its dual is k-uniform.

B.2 f-Dispersers

The following definition is another generalization of disperser graphs.

Definition 27 *Let $G = (V^1, V^2, E)$ be a bipartite graph. An Independent-Set $I \subseteq V$ is called a $[\delta_1, \delta_2]$-Independent-Set if it satisfies:*

$$|I \cap V^1| = \delta_1 V^1 \ , \ |I \cap V^2| = \delta_2 V^2$$

Definition 28 *Let $G = (V^1, V^2, E)$ be a bipartite graph. G is called a (δ_1, δ_2)-disperser if it does not contain any $[\delta_1, \delta_2]$-Independent-Set nor any $[\delta_2, \delta_1]$-Independent-Set.*

Definition 29 *Let $G = (V^1, V^2, E)$ be a bipartite graph. G is called an f-disperser if it is a (δ_1, δ_2)-disperser for every $\delta_1 > \delta_2$ so that $\delta_2 = f(\delta_1 + \delta_2)$. G is called an $(f + \varepsilon)$-disperser if it is a g-disperser for a function g that satisfies*

$$\forall x \ . \ g(x) \leq f(x) + \varepsilon$$

Definition 30 *For every $k \geq 3$ let $\{\Omega_k : [0,1] \mapsto [0,1]\}$ be a family of functions with the following properties: Ω_k achieves a maximum in the range $(0,1)$ at the point $\alpha_k \in (\frac{1}{k}, 1)$. Ω_k is monotone decreasing in the range $(\alpha_k, 1)$. Define the line:*

$$l_k(x) = \Omega_k(\alpha_k) \cdot \frac{1-x}{1-\alpha_k}$$

The function Ω_k is smaller or equal to l_k in the range $(\alpha_k, 1)$, formaly:

$$\forall x \in (\alpha_k, 1) \ . \ \Omega_k(x) \leq l_k(x)$$

Furthermore, the constant α_k satisfies:

$$\alpha_3 \leq 0.919, \alpha_4 \leq 0.842, \alpha_5 \leq 0.778$$

$$\alpha_k < \frac{4 \ln k}{k}$$

Lemma 31 *For every* $\varepsilon > 0$ *and* $N > 0$ *there exists a k-regular k-strongly-edge-colorable* $(\Omega_k+\varepsilon)$*-disperser* $G = (V^1, V^2, E)$*, for which*

$$|V^1| = |V^2| = T \cdot N$$

Where T *is a sufficiently large constant that is independent of* N*.*

Proof: See Appendix C. ∎

C k-Partite k-Uniform Balanced δ-Hyper-Dispersers

In this section, we prove lemma 13, by demonstrating the existence of its dual graph. Namely, we prove the existence k-Partite k-uniform balanced $\frac{1}{k^2}$-hyper-disperser that is $O(k \ln k)$-regular and $3k \ln k$ strongly-edge-colorable. As stated before, these are generalizations of disperser graphs. In addition, we provide an explanation why these are the best (up to a constant) parameters for a hyper-disperser one can hope to achieve.

Construction. First, notice that if a hyper-graph is a δ disperser according to definition 28, then for every set of vertices $U \subseteq V$, if $U^i = U \cap V^i$ and $U^j = \max_i\{U^i\}$ it holds that:

$$\sum_{i \neq j} U^i \leq \delta V$$

We now turn to the proof itself:
The proof is by standard *probabilistic method* (see [AS00]). We construct a random k-partite k-uniform d-regular d-edge-colorable hyper-graph and show that if $d = 2k \ln k$ then with high probability the hyper-graph is a $\frac{3}{k^2}$-hyper-disperser. This implies the existence of such graphs.

Construct a hyper-graph H as follows:

1. Begin with a k-partite graph with no edges. The vertices are: $V = \cup_i V^i$. Denote $V^i = \{v[i, j] | j \in [n]\}$. We call each part V^i a *column*.

2. Define d sets of k-edges, one for each color. The i'th set of edges is denoted $E_i = \{e[i, j] | j \in [n]\}$, and defined as following:

   - Define $k - 1$ random permutations: $\Pi^i_j \in_R S_n | j \in \{2, ..., k\}$ (chosen uniformly from the set of all permutations on n items).
   - In order to define an edge, simply take the an item from the first column, and follow the first permutation to the second column, the second permutation to the third column and so on. Continue this process and define the overall choice as a k-edge. Formally, the j'th edge in the i'th set of k-edges is:

$$e[i, j] = \{\ v[1, j],\ v[2, \Pi^i_2(j)]\ , ...,\ v[k, \Pi^i_k(j)]\ \}$$

Notice that each set of edges E_i covers every vertex exactly once. Therefore, H is edge-colorable with d colors.

Proof:[**of Lemma 13**] We proceed to show that if $d = \Omega(k \ln k)$, then the graph H is a $\frac{1}{k^2}$-hyper-disperser with high probability. Denote by P the probability that the random hyper-graph

H (from the above construction) is not a $\frac{1}{k^2}$-hyper-disperser. Let $\mathcal{U}$ be the family of all subsets $U \subseteq V$ of interest, namely,

$$\mathcal{U} = \{U | U \subseteq V, |U| = \frac{2n}{k}, U \cap V^1 = \frac{n}{k}\}$$

(note that if a set U has no two vertices in one edge, so does any subset of U, hence it suffices to check H for all subsets $U \in \mathcal{U}$).

Denote by $P[U]$ the probability (over H) that **no** two vertices of U share an edge. By union bound,

$$\begin{aligned} P &= \Pr[\exists U \in \mathcal{U}, \text{and no two vertices of } U \text{ share an edge}] \\ &\leq \sum_{U \in \mathcal{U}} P[U] \leq \binom{(k-1)n}{\frac{n}{k}} \binom{n}{\frac{n}{k}} P[\hat{U}] \end{aligned} \quad (2)$$

where $\hat{U} \in \mathcal{U}$ is the set which maximizes $P[\hat{U}]$.

We next bound $P[\hat{U}]$. Let $U^i = \hat{U} \cap V^i$. Let $A_{i,j}$ be the event that there is an edge that contains vertices both from U^i and from U^j, and define some arbitrary linear order on the set of (unordered) couples $(i,j) \in [k] \times [k]$. Then

$$P[\hat{U}] = \Pr[\bigcap_{(i,j)} \neg A_{i,j}] = \prod_{i,j} \Pr\left[\neg A_{i,j} | \bigcap_{(i',j')<(i,j)} \neg A_{i',j'}\right]$$

We know that $\Pr[A_{i,j} | \bigcap_{(i',j')<(i,j)} \neg A_{i',j'}] \geq \Pr[A_{i,j}]$ as no collisions between $U_{i'}$ and $U_{j'}$ means more occupied hyper edges. Hence we have

$$P[\hat{U}] \leq \prod_{i,j} \Pr[\neg A_{i,j}] \leq \prod_{i,j} [1 - \frac{U^i}{n}]^{dU^j}$$

As each vertex from U^i is of degree d, and collide with a vertex vertices from U^j with probability at least $\frac{U^j}{n}$ for each random edge. Therefore:

$$P[\hat{U}] \leq \prod_{i,j} e^{-\frac{dU^i U^j}{n}} = e^{-\frac{d}{n} \sum_{i=1}^{k} U^i \sum_{j=i+1}^{k} U^j}$$

Under the constraint that $\hat{U} \in \mathcal{U}$ the sum $\sum_{i=1}^{k} U^i \sum_{j=i+1}^{k} U^j$ is minimized for $U^1 = U^2 = \frac{n}{k}$ hence

$$P[\hat{U}] \leq e^{-\frac{dn}{k^2}}$$

Therefore by equation 2,

$$P \leq ((ek^2)^{\frac{n}{k}} (ek)^{\frac{n}{k}}) e^{-\frac{dn}{k^2}} \leq (e^2 k^3)^{\frac{n}{k}} e^{-\frac{dn}{k^2}}$$

Thus any d which guarantees that $(e^2k^3)^{\frac{n}{k}} e^{-\frac{dn}{k^2}} \ll 1$ suffices; hence $\frac{n}{k} \ln e^2 k^3 < \frac{dn}{k^2}$ and we have $d > 3k \ln k + 2k$

■

We now turn to see why the hyper-disperser we built above has optimal parameters. We base our observation on a theorem by Ta-Shma et al from [RTS00]:

Lemma 32 (Ta-Shma et al) *Every bipartite d-regular* $(\frac{1}{k}, \frac{1}{k})$*-disperser must satisfy* $d = \Omega(k \log k)$.

Proposition 33 *Every k-partite k-uniform d-regular* $\frac{1}{k^2}$*-hyper-disperser must satisfy* $d = \Omega(k \log k)$.

Proof: We prove that in case there exists a hyper-graph which satisfies $d = o(k \log k)$, then there exists a bipartite $o(k \log k)$-regular $(\frac{1}{k}, \frac{1}{k})$-disperser, in contrast to theorem 44.
And indeed, we can transform a k-partite k-uniform d-regular $\frac{1}{k^2}$-hyper-disperser $H = (V^1, V^2, ..., V^k, E_H)$ into a bipartite d-regular $(\frac{1}{k}, \frac{1}{k})$-disperser $G = (U^1, U^2, E_G)$ in the following way:
Let:

$$U^1 = V^1$$

$$U^2 = V^2$$

$$E_G = \{(v_1, v_2) | (v_1, v_2, v_3, ..., v_k) \in E_H, v_i \in V^i\}$$

Obviously G is a bipartite d-regular graph. In addition, suppose two sets of fractional sizes:

$$S_1 = \frac{1}{k} U_1, S_2 = \frac{1}{k} U_2$$

do not collide in G. Then these sets are of fractional size $\frac{1}{k} \cdot \frac{1}{k-1} \geq \frac{1}{k^2}$ (for large k) in H, thus contradicting the fact that H is a $\frac{1}{k^2}$-hyper-disperser. ∎

D Generalized Dispersers

In this section, we prove the existence of k-regular $(\Omega_k + \varepsilon)$-dispersers that are k strongly-edge-colorable, as stated in lemma 31. In addition, we provide an explanation why these are the best (up to a constant) parameters for an f-disperser one can hope to achieve.

Proof:[**of Lemma 31**] The proof is by the *probabilistic method* (see [AS00]). We construct a random k-regular k-edge-colorable bipartite graph on N vertices and show that with high probability the graph is an $(\Omega_k + \varepsilon)$-disperser (see definitions 29 and 30) for *some* $\varepsilon > 0$. This implies the existence of such graphs. Then we shall prove that $(\Omega_k + \varepsilon)$-disperser for *every* $\varepsilon > 0$ can be constructed using $T \cdot N$ vertices (for some constant $T > 0$ which is independent of N).

Construct a bipartite graph $G = (V^1, V^2, E)$ as follows:
item Let V^1, V^2 be the sets of vertices. Denote:

$$V^1 = \{v^1[j] | j \in [N]\}$$

$$V^2 = \{v^2[j] | j \in [N]\}$$

Define k sets of edges, one for each color:

$$E = \cup_{i \in [k]} E_i$$

The i'th set of edges E_i is denoted:

$$E_i = \{e^i_j | j \in [N]\}$$

and defined as following: Let $\Pi^i \in_R S_N$ be a uniformly chosen random permutation. Each edge e^i_j is defined using Π^i. Namely, the edge e^i_j is:

$$e^i_j = (\ v^1[j],\ v^2[\Pi^i(j)]\ )$$

Let $\Omega_k : [0,1] \mapsto [0,1]$ be the following function:

$$\Omega_k(\delta) \equiv \min \begin{cases} \frac{1}{2}\delta \\ \\ \eta \text{ s.t. } D(\delta - \eta, \eta, k) = 1 - \zeta \end{cases}$$

Where $\zeta = 10^{-3}$ (an arbitrary choice, any constant close to 1 shall do) and:

$$D(\delta, \eta, k) \quad \equiv \quad \frac{(1-\eta)^{(k-1)(1-\eta)} \cdot (1-\delta)^{(k-1)(1-\delta)}}{(1-\eta-\delta)^{k(1-\eta-\delta)} \cdot \delta^\delta \cdot \eta^\eta}$$

Intuitively, the function Ω_k regards every input δ as the fractional size of an independent set in the graph G. For this fractional size, Ω_k outputs the maximum feasible minority fraction. That is, suppose that G contains an $[\delta_1, \delta_2]$-independent set of fractional size $\delta = \delta_1 + \delta_2$ and $\delta_1 \leq \delta_2$. Then δ_1 must satisfy $\delta_1 \leq \Omega(\delta)$.
The expression of Ω_k contains a minimum between $\frac{1}{2}\delta$ (which is an obvious bound on the minority) and the value η s.t. $D(\delta - \eta, \eta, k) = 1 - \zeta$. The expression $D(\delta - \eta, \eta, k)^N$ represents an approximation of the probability that G contains a $[\delta - \eta, \eta]$-independent set.
We next prove that the graph G constructed above is an Ω_k-disperser.

Lemma 34 *With high probability, G is an $(\Omega_k + \varepsilon)$-disperser.*

Proof: We calculate the probability that G contains an $[\eta, \delta]$-independent set or an $[\eta, \delta]$-independent set for *any* $\eta \leq \delta$ so that $\eta \geq \Omega_k(\eta + \delta) + \varepsilon$. Denote this probability by $\hat{P}$. Denote the probability that G contains an $[\eta, \delta]$-independent set or an $[\eta, \delta]$-independent set for *a specific* $\eta \leq \delta, \eta = \Omega_k(\eta + \delta) + \varepsilon$ by P_η. According to the union bound:

$$\hat{P} \leq \sum_{\eta N=1}^{N} P_\eta \leq N \cdot P_\eta$$

Let $U_i^1 \subseteq V^1$ be a set of vertices of cardinality ηN and $U_j^2 \subseteq V^2$ a set of cardinality δN so that $\eta = \Omega_k(\eta + \delta) + \varepsilon$. Denote by $A_{i,j}$ the event that $[U_1, U_2]$ is an independent set in G. According to the construction of G the probability of this event is:

$$P[A_{i,j}] = \left(\binom{(1-\eta)N}{\delta N} \cdot \binom{N}{\delta N}^{-1} \right)^k$$

As this is the probability that k permutations assign each vertex of the set U_j^2 a vertex in the set $V^1 \setminus U_i^1$. Therefore, according to the union bound the probability that G contains an $[\eta, \delta]$-independent set or a $[\delta, \eta]$-independent set for a specific η is bound by:

$$\begin{aligned}
P_\eta &= P[\cup_{i,j} A_{i,j} \ \cup \ \cup_{j,i} A_{j,i}] \\
&\leq 2 \sum_{i,j} P[A_{i,j}] \\
&\leq 2 \binom{N}{\delta N} \binom{N}{\eta N} \cdot \left(\binom{(1-\eta)N}{\delta N} \cdot \binom{N}{\delta N}^{-1} \right)^k \\
&= 2 \binom{N}{\delta N \ \ \eta N} \cdot \left(\binom{(1-\eta)N}{\delta N} \cdot \binom{N}{\delta N}^{-1} \right)^{k-1}
\end{aligned}$$

Therefore:

$$\hat{P} \leq 2N \cdot \binom{N}{\delta N \ \ \eta N} \cdot \left(\binom{(1-\eta)N}{\delta N} \cdot \binom{N}{\delta N}^{-1} \right)^{k-1}$$

Claim 35

$$\sqrt[N]{\hat{P}} \leq \sqrt[N]{20} \cdot D(\delta, \eta, k)$$

Proof: We break up $\sqrt[N]{\hat{P}}$ into expressions:

$$\sqrt[N]{\hat{P}} \leq \sqrt[N]{2N} \cdot C(\delta, \eta) \cdot A^k(\delta, \eta)$$

Where:

$$A(\delta, \eta) \equiv \sqrt[N]{\binom{(1-\eta)N}{\delta N} \cdot \binom{N}{\delta N}^{-1}}$$

$$C(\delta, \eta) \equiv \sqrt[N]{\binom{N}{\delta N \ \ \eta N}}$$

By the Stirling approximation, which states that $n! = \sqrt{2\pi n}(\frac{n}{e})^n(1+\frac{1}{12n}+O(n^{-2}))$:

$$\sqrt{2\pi\alpha n}\left(\frac{\alpha n}{e}\right)^{\alpha n}\left(1+\frac{1}{20\alpha n}\right) \le (\alpha n)! \le \sqrt{2\pi\alpha n}\left(\frac{\alpha n}{e}\right)^{\alpha n}(1+\frac{1}{\alpha n})$$

We get:

$$\begin{aligned}
A^N(\delta,\eta) &= \frac{((1-\eta)N)!\cdot((1-\delta)N)!}{((1-\eta-\delta)N)!N!} \\
&\le \frac{\left(\frac{(1-\eta)N}{e}\right)^{(1-\eta)N}\left(\frac{(1-\delta)N}{e}\right)^{(1-\delta)N}\sqrt{4\pi^2N^2(1-\delta)(1-\eta)}\left(1+\frac{1}{(1-\delta)N}\right)\left(1+\frac{1}{(1-\eta)N}\right)}{\left(\frac{(1-\eta-\delta)N}{e}\right)^{(1-\eta-\delta)N}(\frac{N}{e})^N\sqrt{4\pi^2N^2(1-\delta-\eta)}\left(1+\frac{1}{20(1-\delta-\eta)N}\right)(1+\frac{1}{20N})} \\
&\le \left[\frac{(\frac{N}{e})^{(1-\eta)+(1-\delta)}}{(\frac{N}{e})^{(1-\eta-\delta)+1}}\cdot\frac{(1-\eta)^{1-\eta}\cdot(1-\delta)^{1-\delta}}{(1-\eta-\delta)^{1-\eta-\delta}}\right]^N\cdot\sqrt{1-\frac{\delta\eta}{1-\delta-\eta}}\cdot\left(\frac{1}{1+\frac{1}{20N}}\right)^2 \\
&\le \left[\frac{(1-\eta)^{1-\eta}\cdot(1-\delta)^{1-\delta}}{(1-\eta-\delta)^{1-\eta-\delta}}\right]^N\cdot\left(\frac{N+1}{N}\right)^2 \\
&\le 4\cdot\left[\frac{(1-\eta)^{1-\eta}\cdot(1-\delta)^{1-\delta}}{(1-\eta-\delta)^{1-\eta-\delta}}\right]^N
\end{aligned}$$

Similar algebra yields:

$$\begin{aligned}
C^N(\delta,\eta) &= \frac{(N)!}{(\delta N)!\cdot(\eta N)!((1-\eta-\delta)N)!} \\
&\le \frac{\sqrt{2\pi N}\left(\frac{N}{e}\right)^N(1+\frac{1}{N})}{\sqrt{2^3\pi^3N^3(1-\eta)(1-\delta)(1-\delta-\eta)}(\frac{\delta N}{e})^{\delta N}\cdot(\frac{\eta N}{e})^{\eta N}\cdot(\frac{(1-\eta-\delta)N}{e})^{(1-\eta-\delta)N}\left(1+\frac{1}{20N}\right)^3} \\
&\le \left(\frac{1}{2\pi N}\right)\cdot\left[\frac{1}{\delta^\delta\cdot\eta^\eta\cdot(1-\eta-\delta)^{1-\eta-\delta}}\right]^N\left(\frac{N+1}{N}\right)^3 \\
&= \frac{4}{\pi N}\cdot\left[\frac{1}{\delta^\delta\cdot\eta^\eta\cdot(1-\eta-\delta)^{1-\eta-\delta}}\right]^N
\end{aligned}$$

Combining both results yields:

$$\sqrt[N]{\hat{P}} \le \sqrt[N]{\frac{32}{\pi}}\cdot D(\delta,\eta,k) < \sqrt[N]{20}\cdot D(\delta,\eta,k)$$

■

Since $\eta \le \delta$, we know that $\eta \le \frac{\eta+\delta}{2}$. Therefore, the fact that $\eta > \Omega_k(\eta+\delta)$ implies $\Omega_k(\eta+\delta) \ne \frac{\delta+\eta}{2}$ and hence $D(\delta,\eta,k) = 1-\zeta$.
Therefore $\sqrt[N]{\hat{P}} < \sqrt[N]{20}(1-\zeta)$. This implies that for a sufficiently large (constant) N, $\hat{P} << 1$. Hence G is an $(\Omega_k+\varepsilon)$-disperser with high probability. Notice that for every $\varepsilon > 0$, taking $\hat{N} = N\cdot T$ for some constant T suffices for G to be an $(\Omega_k+\varepsilon)$-disperser with high probability.
■

Thus, we have proved that the random graph G is an $(\Omega_k+\varepsilon)$-disperser with high probability. It remains to prove the properties of the function Ω_k according to definition 30.

Proposition 36 *There exists a constant α_k for which $\Omega_k(\alpha_k)$ is a strict maximum for the function Ω_k in the range $(0,1)$. The constant α_k satisfies:*

$$\alpha_3 \leq 0.919, \alpha_4 \leq 0.842, \alpha_5 \leq 0.778$$

$$\alpha_k < \frac{4\ln k}{k}$$

Proof: Define as α_k the single constant in the range $0 < \alpha_k < 1$ for which:

$$D(\frac{\alpha_k}{2}, \frac{\alpha_k}{2}, k) = 1 - \zeta$$

We first explain why such a constant exists. Define the function $D_k(x)$ as:

$$D_k(x) := D(x, x, k) = \frac{(1-x)^{2(k-1)(1-x)}}{(1-2x)^{k(1-2x)} \cdot x^{2x}}$$

Define the natural logarithm of $D_k(x)$ as:

$$g_k(x) := \ln D_k(x) = 2(k-1)(1-x)\ln(1-x) - k(1-2x)\ln(1-2x) - 2x\ln x$$

The first and second derivatives of g_k are:

$$g_k'(x) = -2(k-1)\ln(1-x) + 2k\ln(1-2x) - 2\ln x$$

$$g_k''(x) = \frac{2(k-1)}{1-x} - \frac{2k}{1-2x} - \frac{2}{x}$$

Notice that the second derivative g_k'' is negative in the range $(0, \frac{1}{2})$. Hence, g_k' is monotone decreasing. In addition, the first derivative satisfies:

$$\lim_{x \to 0^+} g_k'(x) = \infty \ , \ \lim_{x \to \frac{1}{2}^-} g_k'(x) = -\infty$$

Therefore, in the range $(0, \frac{1}{2})$ the function g_k is first increasing, then attains a maximum and then decreases. Because the natural logarithm is a monotone function, the function $D_k(x)$ behaves the same way. In addition, D_k accepts the values:

$$\lim_{x \mapsto 0^+} D_k(x) = 1^+, \lim_{x \mapsto \frac{1}{2}^-} D_k(0.5) = 0+$$

Hence, in the range $(0, \frac{1}{2})$, the function $D_k(x)$ begins from the value of 1, raises to the maximum and from the maximum decreases monotonically to zero. Therefore, the function $D(\frac{x}{2}, \frac{x}{2}, k)$ behaves the same way in the range $(0,1)$. In particular, there is a unique point $\alpha_k \in (0,1)$ so that $D(\frac{\alpha_k}{2}, \frac{\alpha_k}{2}, k) = 1 - \zeta$. In addition, the function D is monotone decreasing in the range $(\alpha_k, 1)$.

Claim 37 *For every $\delta > \frac{1}{2k}$, the function $D(\delta, \delta, k)$ satisfies:*

$$\forall \delta \in (\tfrac{1}{2k}, \tfrac{1}{2}) \ . \ \forall \alpha \in [0,1] \ . \ D(\delta, \delta, k) \leq D(2\delta \cdot \alpha, 2\delta \cdot (1-\alpha), k)$$

Proof: First notice the equivalent definition the property above:

$$\forall \delta+\eta \in (\frac{1}{k},1)\ .\ D(\delta,\eta,k) \geq D(\frac{\delta+\eta}{2},\frac{\delta+\eta}{2},k)$$

According to the definition of the function D:

$$D(\delta,\eta,k) = \frac{(1-\eta)^{(k-1)(1-\eta)} \cdot (1-\delta)^{(k-1)(1-\delta)}}{(1-\eta-\delta)^{k(1-\eta-\delta)} \cdot \delta^{\delta} \cdot \eta^{\eta}}$$

Because part of the denominator is the same for any constant sum $\delta+\eta$, it is enough to see that the function $f(x) = \frac{(1-x)^{(k-1)(1-x)}}{x^x}$ is multiplicative convex in the range $\frac{1}{k} < x < 1$. That is, for $x_1 + x_2 \in (\frac{1}{k}, 1)$:

$$f(x_1) \cdot f(x_2) = \frac{(1-x_1)^{(k-1)(1-x_1)}}{x_1^{x_1}} \cdot \frac{(1-x_2)^{(k-1)(1-x_2)}}{x_2^{x_2}} \geq \frac{(1-\frac{x_1+x_2}{2})^{(k-1)(2-x_1-x_2)}}{\left(\frac{x_1+x_2}{2}\right)^{x_1+x_2}} = f^2(\frac{x_1+x_2}{2})$$

This of course holds if the function

$$g(x) := \log f(x) = (k-1)(1-x)\log(1-x) - x\log x$$

is convex in the specified range.
And indeed function $g(x)$ is convex in the range $(\frac{1}{k}, 1)$ as can be seen by the taking the second derivative:

$$g'' = \frac{xk-1}{x(1-x)}$$

Which is positive in this range.

■

Corollary 38 *For every $\eta \geq \delta > \frac{1}{2k}$, the function $D(\delta,\eta,k)$ satisfies:*

$$\forall \varepsilon \in (0, \tfrac{\eta+\delta}{2})\ .\ \forall \alpha \in [0,1]\ .\ D(\delta+\varepsilon, \eta-\varepsilon, k) \leq D(\delta,\eta,k)$$

Proof: Notice that in the proof of claim 37, the function $g(x)$ was shown to be convex. The corollary follows from the following fact about the convex function g:

$$\forall \varepsilon \in \left[0, \frac{x_1+x_2}{2}\right]\ .\ g(x_1+\varepsilon) + g(x_2-\varepsilon) \leq g(x_1) + g(x_2)$$

(This know fact in convex theory can be proven using the Tailor series of g). ■

Definition 39 *Define the function:*

$$f_k^{\delta}(x) := D(\delta, x, k)$$

Claim 40 *For $x \geq \delta > \frac{\alpha_k}{2}$ the function $f_k^{\delta}(x)$ is monotone strictly decreasing in the range $x \in (\delta, 1-\delta)$.*

Proof: Define the natural logarithm of $f_k^\delta(x)$ as:

$$q_k^\delta(x) := \ln f_k^\delta(x) = w(\delta)\cdot[2(k-1)(1-x)\ln(1-x) - k(1-x-\delta)\ln(1-x-\delta) - 2x\ln x]$$

Where $w(\delta)$ is a positive expression that depends only on δ. The first and second derivatives of q_k^δ are:

$$q_k'(x) = -2(k-1)\ln(1-x) + 2k\ln(1-x-\delta) - 2\ln x$$

$$q_k''(x) = \frac{2(k-1)}{1-x} - \frac{2k}{1-x-\delta} - \frac{2}{x}$$

Notice that the second derivative g_k'' is negative in the range $(0, 1-\delta)$. Hence, g_k' is monotone decreasing. In addition, the first derivative satisfies:

$$\lim_{x\to 0^+} q_k'(x) = \infty\ ,\quad \lim_{x\to(1-\delta)^-} q_k'(x) = -\infty$$

Therefore, in the range $(0, 1-\delta)$ the function q_k is first increasing, then attains a maximum and then decreases. Because the natural logarithm is a monotone function, the function $f_k^\delta(x)$ behaves the same way. In addition, notice that:

$$\lim_{x\to 0^+} f_k^\delta(x) \geq 1\ ,\ f_k^\delta(\delta) < 1$$

Therefore, the function f_k^δ is monotone strictly decreasing in the range $x \in (\delta, 1-\delta)$. ■

Corollary 41 *For every $\delta \in (\alpha_k, 1)$, the function Ω_k is monotone strictly decreasing.*

Proof: Let $\delta > \alpha_k$ and $\Omega_k(\delta) = x$. Let $\varepsilon > 0$ and $\Omega_k(\delta+\varepsilon) = y$. Suppose that $y \geq x$, then according to corollary 38 and claim 40:

$$D(y, \delta+\varepsilon-y, k) < D(x, \delta+\varepsilon-x, k) < D(x, \delta-x, k) = 1-\zeta$$

In contrast to the assumption that $\Omega_k(\delta+\varepsilon) = y$ which imposes $D(y, \delta+\varepsilon-y, k) = 1-\zeta$. Thus it must be that $y < x$. ■

We proceed to show that $\Omega(\alpha_k)$ is a strict maximum for Ω_k in the range $(0, 1)$.
In the range $0 < x < \alpha_k$, according to the definition of Ω_k, it holds that:

$$\forall x \in (0, \alpha_k)\ .\ \Omega_k(x) \leq \frac{x}{2} < \frac{a_k}{2} = \Omega(\alpha_k)$$

In the range $\alpha_k < x < 1$, $\Omega(\alpha_k)$ is the strict maximum according to corollary 41. Hence, $\Omega(\alpha_k)$ is indeed a strict maximum for Ω_k in the entire range $(0, 1)$.

We proceed to prove the bounds on α_k. For that, it suffices to show that:

$$D(\frac{\alpha_k}{2}, \frac{\alpha_k}{2}, k) < 1-\zeta$$

For the values $k \in [3, 6]$ this is numerically verified. For a general k paramter, we first define the following function which bounds $D(\delta, \eta, k)$:

$$U(\delta, \eta, k) \ := \ \frac{e^{\delta+\eta}e^{-(k-1)\delta\eta}}{\delta^\delta\cdot\eta^\eta}$$

Claim 42 *The function $D(\delta, \eta, k)$ is bound by:*

$$D(\delta, \eta, k) \quad \leq \quad U(\delta, \eta, k)$$

Proof: The function A is bound by:

$$A(\delta, \eta) \leq \sqrt[N]{(1-\eta)^{\delta N}} \leq e^{-\eta\delta}$$

And the function $C(\delta, \eta)$ is bound by (using the inequality $\binom{N}{k} \leq (\frac{eN}{k})^k$):

$$\begin{aligned} C(\delta, \eta) &= \sqrt[N]{\binom{N}{\delta N \quad \eta N}} \\ &\leq \left(\frac{eN}{\delta N}\right)^{\delta} \cdot \left(\frac{eN}{\eta N}\right)^{\eta} \\ &= \left(\frac{e}{\delta}\right)^{\delta} \cdot \left(\frac{e}{\eta}\right)^{\eta} \end{aligned}$$

And the bound on D follows. ■

And now using claim 42:

$$\begin{aligned} D(\frac{1}{k}, \frac{1}{k}, 4k \ln k) \leq U(\frac{1}{k}, \frac{1}{k}, 4k \ln k) &= \frac{e^{\frac{2}{k}} e^{-(4k \ln k - 1)\frac{1}{k^2}}}{\frac{1}{k}^{\frac{2}{k}}} \\ &= (ek)^{\frac{2}{k}} e^{-(4k \ln k - 1)\frac{1}{k^2}} \\ &= e^{\frac{2}{k}(1+\ln k)} e^{-(4k \ln k - 1)\frac{1}{k^2}} \\ &= e^{\frac{2}{k}[1+\ln k - 2\ln k + \frac{1}{2k}]} \\ &\leq e^{\frac{2}{k}[-\ln k + 2]} \\ &\leq e^{-\frac{\ln k}{k}} \\ &< 1 - \zeta \end{aligned}$$

■

Proposition 43 *The function Ω_k is bounded by the line $l_k(x)$ in the range $(\alpha_k, 1)$. Formally:*

$$\forall x \in (\alpha_k, 1) \ . \ l_k(x) - \Omega_k(x) > 0$$

Where:

$$l_k(x) = \frac{\alpha_k}{2(1-\alpha_k)}(1-x)$$

Proof: Since the function $D(\delta, x)$ is monotone decreasing (see claim 40), it suffices to show that for the function D, the following holds:

$$\forall x \in (\alpha_k, 1) \ . \ D(x - l_k(x), l_k(x)) \geq 1$$

In order to simplify expressions, denote by:

$$c = \frac{\alpha_k}{2(1-\alpha_k)}$$

The function $D(x - l_k(x), l_k(x))$ is a function of a single variable. Denote it by:

$$\begin{aligned} Y(x) := D(x - l_k(x), l_k(x), k) &= D(x - c(1-x), c(1-x), k) \\ &= D((1+c)x - c, c(1-x), k) \\ &= \frac{(1+c)(1-x)^{(k-1)(1-x)(1+c)} \cdot (1-c+cx)^{(k-1)(1-c+cx)}}{(1-x)^{k(1-x)} \cdot (c(1-x))^{c(1-x)} \cdot ((1+c)x - c))^{(1+c)x-c}} \end{aligned}$$

The function $Y(x)$ is continues and differentiable function in the range $(0, 1)$, as it is the multiplication of continues and differentiable functions in this range. The first derivative of this function is:

$$Y'(x) = Y(x) \cdot g(x)$$

Where:

$$g(x) := c(k-1)\ln(1-c+cx) - (1+c)(k-1)\ln(1-x-cx+c) + k\ln(1-x) - (1+c)\ln(x+cx-c) + c\ln(c-cx)$$

Notice that $Y'(x)$ is also a continues and differentiable function in the range $(\alpha_k, 1)$, since the argument of the logarithms is always in this range. To show that, consider $\ln(1 - c + cx)$ and $\ln(x + cx - c)$ (for the others this is obvious). The argument of these logarithms is zero outside the range as:

$$1 - c + cx = 0 \Rightarrow x = -\frac{1-c}{c} < 0$$

$$x + cx - c = 0 \Rightarrow x = \frac{c}{1+c} = \frac{\alpha_k}{2-\alpha_k} < \alpha_k$$

The derivative of $g(x)$ is:

$$g'(x) = \frac{[c(1+c)(k-2)]x + 2c^2 - 1}{(1-x)(x+cx-c)(1-c+cx)}$$

This function is also continues in the range $(\alpha_k, 1)$ as the denominator doesn't evaluate to zero in the range (as shown for the arguments of the logarithms of $g(x)$).
Notice that the nominator evaluates to zero exactly once on the real line. The point where this happens is:

$$x = \frac{1-2c^2}{c(1+c)(k-2)} < \frac{1-c}{c(k-2)} = \frac{2-3\alpha_k}{\alpha_k(k-2)}$$

We show that this point is never in the range $(\alpha_k, 1)$. For a large k parameter, the constant c approaches zero, and hence $2c^2 - 1$ is negative. This implies that the point x is negative, and hence not in the range. For small k values, one can substitute the value α_k and see it is larger then $\frac{2-3\alpha_k}{\alpha_k(k-2)}$.
Hence $g'(x)$ is always non-zero in the range $(\alpha_k, 1)$.

Therefore, the function $g(x)$ can accept at most once zero in the range (other wise there would be a maximum or a minimum). Therefore, the function $Y'(x)$ can evaluate to zero only once in the range, as it is the multiplication of $g(x)$ and a positive function.
Hence, the function Y, which is equal to 1 in $\alpha_k - \varepsilon$ and 1, cannot cut the line $l(x) = 1 - \zeta$ inside the range. Therefore, $Y(x)$ is either all below or all above the line $l(x) = 1$. Since:

$$\lim_{x \to 1^-} Y(x) = 1^+$$

Thus, in the range $(\alpha_k, 1)$, the function $Y(x)$ is all above the line $l(x) = 1$. ∎
Combining propositions 36 and 43, completes the properties of Ω_k to be proved.

∎

We now turn to see why the Ω_k-disperser we built above has optimal asymptotic parameters. We base our observation on a theorem by Ta-Shma et al from [RTS00]:

Lemma 44 (Ta-Shma et al) *Every bipartite d-regular $(\frac{1}{k}, \frac{1}{k})$-disperser must satisfy $d = \Omega(k \log k)$.*

Notice that an Ω_k-disperser is in particular a (α_k, α_k)-disperser. As $\alpha_k = O(\frac{\log k}{k})$ and the Ω_k-disperser is k-regular, the optimality follows from Ta-Shma's lemma.

E Reduction from k-IS to k-DM

This appendix includes a brief description of the reduction from k-DM to k-IS, that proves the following proposition:

Proposition 45 *If k-IS is NP-hard to approximate to within certain factors, then (k+1)-DM is NP-hard to approximate within the same factors.*

The reduction:

1. Let $G = (V, E)$ be a graph instance of k-IS. First, replace the role of the edges and vertices to obtain a hyper-graph $H = (E, V)$. Notice that a matching in H is an Independent-Set in G. In addition, because G is of bounded degree k, the edges of H are of cardinality at most k.

2. Vising's theorem states that any graph of maximum degree k can be edge-colored with $k + 1$ colors so that no vertex participates in two edges of the same color. Therefore, H is strongly vertex colorable with $k + 1$ colors, and thus is $(k + 1)$-partite.

3. In order for the hyper-graph H to be $k+1$ uniform, more "dummy" vertices must be added to the graph.

4. As for making the hyper-graph H balanced, we make $k+1$ copies of the current H and join them to s single hyper-graph so that there is an equal number of vertices for each color.

The reduction hereby shows that inapproximability results for k-IS apply to $(k + 1)$-DM as well (and thus also to $(k + 1)$-Set-Packing).